

**18<sup>th</sup> ICCRTS**  
**“C2 in Underdeveloped, Degraded and Denied Operational Environments”**

**Title: Accelerating Exploitation of Low-grade Intelligence  
Through Semantic Text Processing of Social Media**

**Topic: Data, Information and Knowledge**

**Sue E. Kase, John Dumer  
U. S. Army Research Laboratory  
Computational Information Sciences Directorate  
Tactical Information Fusion Branch  
Building 321 RDRL-CII-C  
Aberdeen Proving Ground, MD 21005**

POC: Sue E. Kase  
U. S. Army Research Laboratory  
Computational Information Sciences Directorate  
Tactical Information Fusion Branch  
Building 321 RDRL-CII-C  
Aberdeen Proving Ground, MD 21005  
[sue.e.kase.civ@mail.mil](mailto:sue.e.kase.civ@mail.mil)  
410-278-9762

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>JUN 2013</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2013 to 00-00-2013</b>	
4. TITLE AND SUBTITLE <b>Accelerating Exploitation of Low-grade Intelligence Through Semantic Text Processing of Social Media</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>U.S. Army Research Laboratory, Computational Information Sciences Directorate, Building 321 RDRL-CII-C, Aberdeen Proving Ground, MD, 21005</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>Presented at the 18th International Command &amp; Control Research &amp; Technology Symposium (ICCRTS) held 19-21 June, 2013 in Alexandria, VA.</b>					
14. ABSTRACT <b>Social media will play a key role in many areas of intelligence operations with the development of knowledge extraction analytics. The data contained within these social networking services present many challenges, but the value obtained from detecting subtle and hidden information exchanges and environmental observations are vast. The analytics that will drive social media knowledge exploitation include advanced text processing technologies allowing multi-faceted concept or frame-based queries within a familiar search engine interface for users who are not expert analysts or who may be operating with limited intelligence resources in disadvantaged information collection situations. In this paper, we demonstrate constructing frame-based semantic search queries using models of concepts and roles extracted from social media content. This approach to searching allows intuitive query formulations requiring less up-front knowledge of the search parameters. We demonstrate this semantic search capability by querying over seven million tweets captured during the Arab Spring uprisings.</b>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>25</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

## **Abstract**

Social media will play a key role in many areas of intelligence operations with the development of knowledge extraction analytics. The data contained within these social networking services present many challenges, but the value obtained from detecting subtle and hidden information exchanges and environmental observations are vast. The analytics that will drive social media knowledge exploitation include advanced text processing technologies allowing multi-faceted concept or frame-based queries within a familiar search engine interface for users who are not expert analysts or who may be operating with limited intelligence resources in disadvantaged information collection situations. In this paper, we demonstrate constructing frame-based semantic search queries using models of concepts and roles extracted from social media content. This approach to searching allows intuitive query formulations requiring less up-front knowledge of the search parameters. We demonstrate this semantic search capability by querying over seven million tweets captured during the Arab Spring uprisings.

## Introduction

Social media is steadily growing in importance as an information source. The brevity of social media content (e.g., 140 characters per tweet) combined with the increasing usage of mobile devices facilitates the sharing of information from anywhere at any time. This phenomenon is giving rise to ever stronger citizen authorship, where millions of social media users share about, discuss, comment on and illustrate recent happenings on a daily basis. The unprecedented volumes of real-time information contributed by large communities of social media users can be researched and searched for sources of information relevant to national security.

There is increasing acceptance within the intelligence community of the potential value of social media, but understandably some reluctance to rank it alongside more traditional clandestine sources. There is no denying that much of the information available from platforms such as Twitter and Facebook is of little use from an intelligence perspective. But there is still a great deal of useful and actionable information available and the challenge for the intelligence community is not only to learn how to extract that information using new tools and techniques, but how to analyze it as a unique and dynamic data source.

The value of social media content as an intelligence source comes to the forefront when traditional methods of intelligence collection such as airborne sensors, photographic devices, infrared detection systems, listening devices, and motion detectors are unavailable, for instance, within a constrained resource allocation environment. Also, there can be challenges and risks involved in direct intelligence collection resulting from unsafe conditions due to military presence or conflicts in progress. Consequently, intelligence collection efforts must sometimes rely on the eyes and ears of the general population. Members of the public go about their daily business observing actions or overhearing information that may be of immeasurable value to security forces. With easier and more widespread access to the Internet, the general population can share their observations and information through communications activities enabled by social media services. There are many social media services to choose from: yahoo groups, google groups, wikis, user forums, Ning, LinkedIn groups, blogs, Facebook, and Twitter—literally, hundreds of ways to share information.

While information collected from social media content may not provide the information superiority of high-grade intelligence, this citizen-authored information could complement various types of high-grade intelligence. Social media sources can provide large amounts of low-grade information that when added together, provide a picture of the political, military, sociological, and infrastructure aspects of an operating environment. The collection of low-grade intelligence is an approach attributed to the famous British counterinsurgency expert General Frank Kitson (Kitson, 1971).

According to a NextGov article (NextGov, 2012) summarizing a panel discussion of open-source intelligence hosted by *Government Executive*, there is potential for social media to take the place of clandestine sources when it comes to intelligence analysis. While this may be overly optimistic, and not necessarily the desired end goal, there is certainly scope for social media to provide useful insights that can either inform or support information gathered from traditional intelligence sources. Referring to the NextGov article, Centrifuge Systems ([www.centrifugesystems.com/news](http://www.centrifugesystems.com/news)) noted that social media could help predict large scale shifts in the cultural, political, and social world. An example would be the impact of social media on shaping geopolitical development during the recent Arab Spring and current Syrian conflict.

This paper introduces a new technology for extracting low-grade intelligence information from large streams of social media communications. The next section overviews this frame-based semantic search technology within a suite of advanced analytics called Contour. Contour, developed by Decisive Analytics Corporation (DAC, <http://www.dac.us/>), offers a range of capabilities to facilitate the search of

social media content. The frame-based ontology used by Contour, called FrameNet, is discussed. The third section demonstrates Contour's multi-level semantic search formulation of Concepts, Roles, and Keywords in the context of a large collection of tweets from the Arab Spring uprisings. Example search queries illustrate how this new semantic search technology can support an analyst's situational awareness, firstly, as to what is occurring within a specific geographical area, and secondly, how this may affect ongoing missions. The final section discusses advantages Contour's semantic search capability offers in the context of a disadvantaged intelligence collection, as well as, needed future areas of development.

## Frame-based Semantic Search

Text commonly contained in social media communication activities is considered "unstructured", shortened, informally written text with embedded hashtags, urls, replies and mentions, and emoticons. In addition to its unstructured nature, the sheer volume of social media communications sent during events of interest is overwhelming and hence difficult to distill for relevant information. In order to utilize social media communications for low-grade intelligence, analysts must rely on automated text processing tools to filter through and organize these large bodies of text into a rapidly-digestible form. Contour's frame-based semantic modeling is one approach to addressing the text processing challenges associated with a large corpus of unstructured text.

The Contour platform imports unstructured text from a variety of sources and then maps the text to an existing ontology of frames (FrameNet, <https://framenet.icsi.berkeley.edu/fndrupal/>) during a process of Semantic Role Labeling (SRL). FrameNet is a structured language model grounded in the theory of Frame Semantics by Charles Fillmore and colleagues (1976). Fillmore's hypothesis is people understand things by performing mental operations on what they already know and such knowledge is describable in terms of information packets called *frames*.

Frame semantic analysis is a technique linguists use to characterize words by their interactions with the words around them. The goal of frame-based text analysis is to identify the words that evoke specific concepts in the reader's mind (*frames*) and to map words and phrases from the text to the *roles* that are related to these concepts. A simple example of the frame concept can be seen in this sentence:

John bought food.

The verb "bought" in this sentence brings to mind, or *evokes* in the reader's mind, a scenario in which goods are exchanged for money. Using the preexisting understanding of a commercial transaction (the schema for the commercial transaction "frame") the reader understands that John plays the role of the buyer and that goods were exchanged. The prototypical commercial transaction scenario acts as a frame that places phrases in the sentence "John bought food" into their semantic roles. Under the frame-based model of semantics, understanding language involves accessing the semantic frames evoked by the text and associating groups of words with the roles defined by the frames.

The FrameNet lexical database of English is both human- and machine-readable based on annotating examples of how words are used in actual text (Ruppenhofer, Ellsworth, Petruck, Johnson, & Scheffczyk, 2006). The FrameNet database is under development at the International Computer Science Institute in Berkeley since 1997 supported primarily by the National Science Foundation. It consists of over 1,000 semantic frames, their associated Frame Elements (FEs) and the words that evoke them. For example, Table 1 shows the *Operate\_vehicle* frame prototypical of a scenario in which someone operates a type of vehicle.

Frame (FE)	Element	Description
Area		General area in which motion takes place
Driver		The being that controls the vehicle as it moves
Goal		Where the moving object ends up
Path		Description of the trajectory of motion
Source		Expression that defines a starting point of motion
Vehicle		The means of conveyance controlled by the driver

Table 1. Several of the Frame Elements (FEs) of the *Operate\_vehicle* frame. FEs are the semantic components of the concept of operating a vehicle.

For each of FrameNet's frame definitions, the database defines a number of words that, when used in the right context, evoke that frame. For example, some of the words that evoke the *Operate\_vehicle* frame are: *drive*, *fly*, *ride*, *row*, and *taxi*.

Contour's semantic search capability employs Semantic Role Labeling (SRL) (Gildea & Jurafsky, 2002) to automatically map words and phrases from the source text into the elements of the frames that are evoked by the ideas in that text. Contour's SRL algorithm uses a supervised training technique to build a probabilistic model that maps constituents in Context Free Grammar (CFG) parse trees to elements in the frames evoked by the text. This FrameNet-based SRL process builds a rich, formal model of meaning from the target text. Figure 1 shows the result of FrameNet-based SRL performed on a simple target sentence. The data structure that contains the mapping between frame concepts and the target text (portion below the arrow in Figure 1) is called the *semantic model* of the text. The semantic model of a target document acts as a high-fidelity representation of the text's meaning.

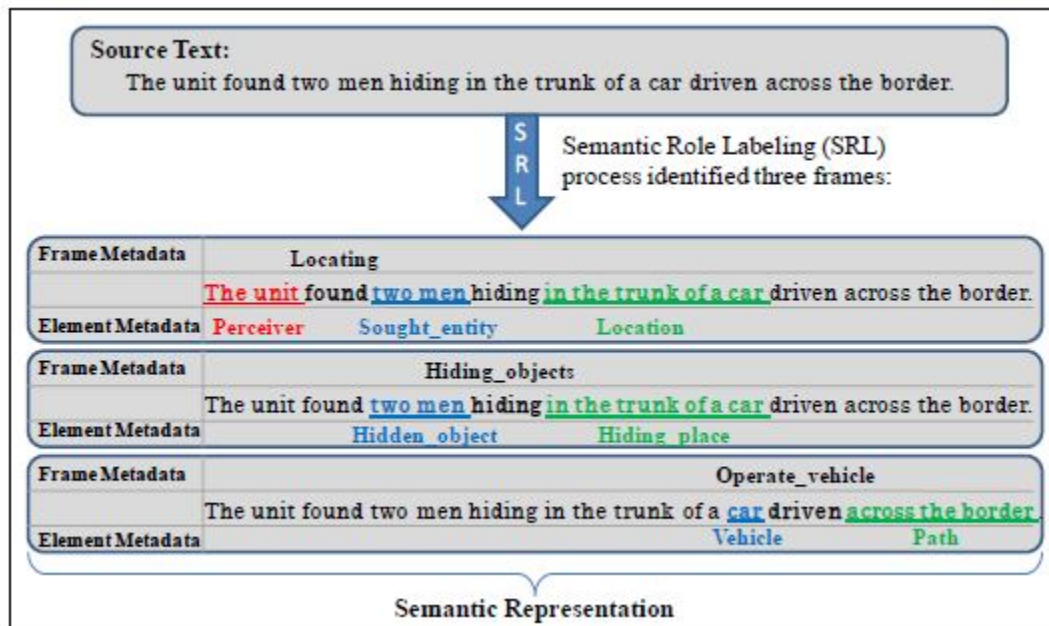


Figure 1. A FrameNet-based Semantic Role Labeling process uncovers three frames in the target sentence. Hiding\_objects (evoked from the word *hiding*), Locating (evoked from the word *found*) and Operate\_vehicle (evoked from the word *driven*). Text from the sentence that fulfills the role of the elements for each frame is labeled with the name of its element. The data structure that maps target text to frames and frame elements is called the *semantic representation*.



In Contour, the FrameNet-based model of the text can be used to construct queries through a series of filtering widgets, or to formulate a more targeted query in a Semantic Search box. The filtering widgets and Semantic Search box are located along the left side of the main window of the user interface (Figure 2). The large subwindow to the right of the filtering and search widgets displays the search results.

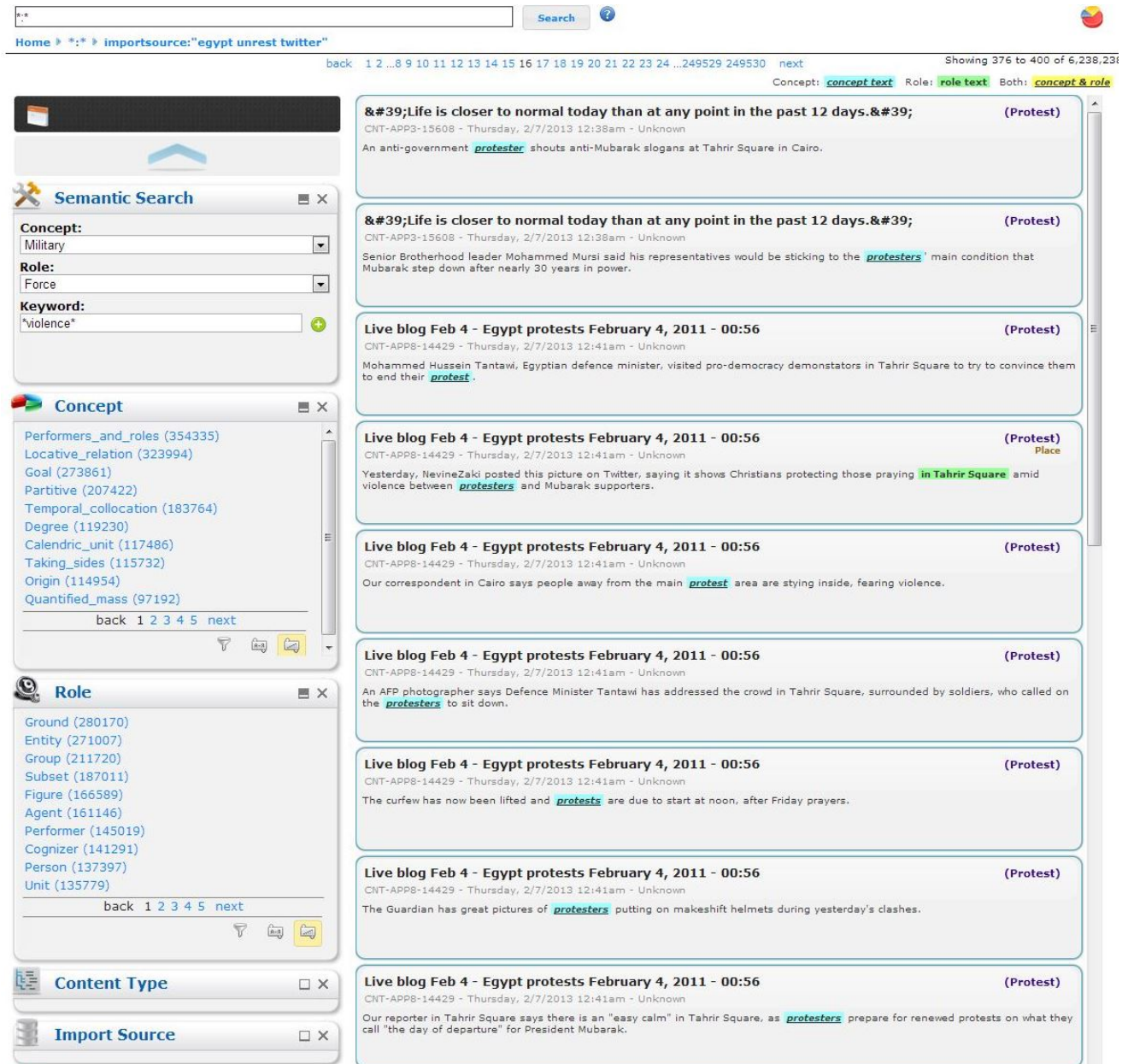


Figure 2. The main window of the user interface to Contour's Semantic Search technology. Filtering widgets and a Semantic Search box are along the left side of the screen. Search results are displayed in the large subwindow on the right and center within the main window.

Figure 3 is an enlarged composite picture of the FrameNet-inspired portions of the user interface with the frame based filtering widgets (left side) and the Semantic Search box (upper right side). On the left side of Figure 3, the FrameNet frame *Killing* has been selected in the Concept filtering widget. The Concept

Killing would identify sentences containing words such as *assassinate*, *bloodshed*, *destroy*, *fatality*, *massacre*, *murder*, *shooting*, and *terminate*. After selecting the Killing Concept, the Role filtering widget will show the decomposition of the Killing frame into its frame elements along with the number of tweets where that Concept and Role are found. For example, using the Killing Concept and Victim Role filters, 4697 tweets were found. On the right side of Figure 3, the Semantic Search box can be used for a targeted search of the Killing frame (in the Concept text box) with frame element Place (in the Role text box) and a Keyword with wildcard characters *\*Cairo\**. This search will return tweets fitting the Killing frame with the frame element of Place matching Cairo. Two example tweets returned by this query are shown on the lower right side of Figure 3 where the blue highlighted text is the Concept text, green text is the Role text, and the yellow text is both the Concept and the Role.

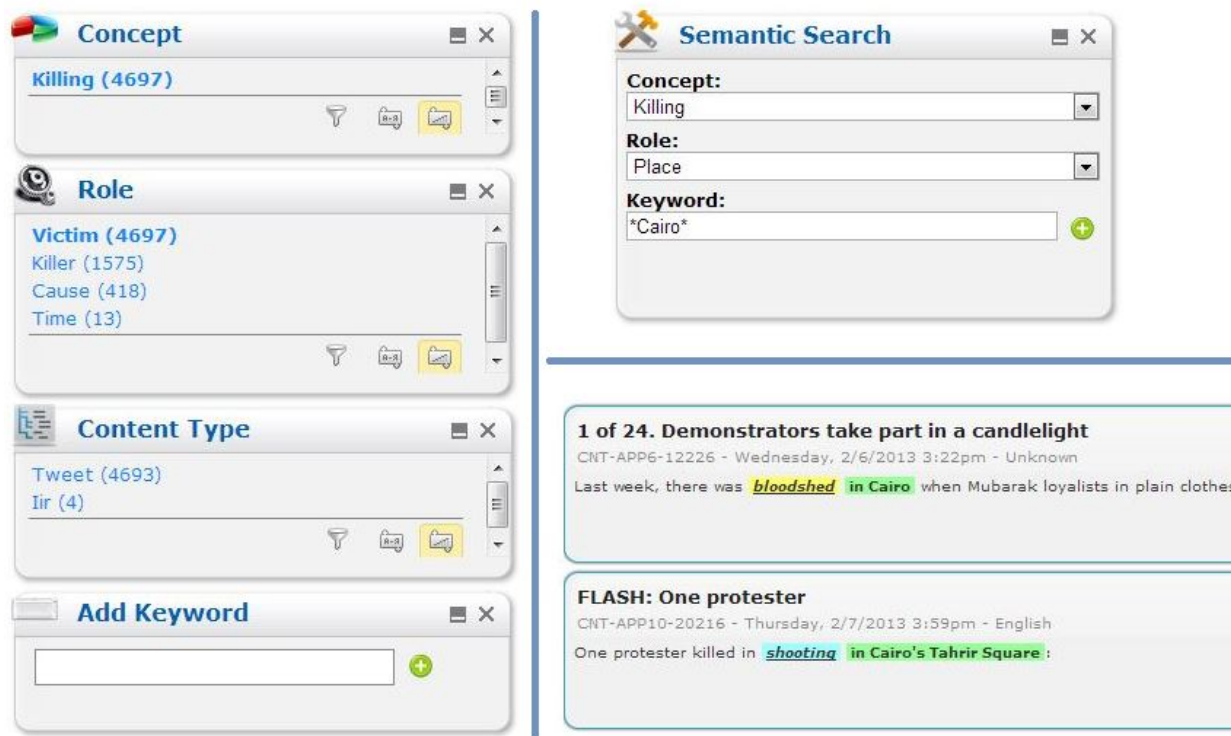


Figure 3. The Concept and Role widgets with the Killing frame selected and associated frame elements displayed in the Role widget (left side). The Semantic Search box containing the Killing Concept, Place Role, and a Keyword representing a specific location *\*Cairo\** (upper right side). Two example tweets resulting from the search criteria selected in the Semantic Search box (lower right side).

By utilizing Contour's frame-based filtering widgets and/or more targeted Semantic Search box a large volume of social media generated communications can be efficiently queried filtering out most of the non-informative text for possible collection of low-grade intelligence information. The next section examines this frame-based semantic search approach in more detail by querying a Twitter dataset of over seven million tweets collected during the height of the Arab Spring protests.

### Use Case: Searching Arab Spring Tweets

Twitter's ever growing daily flow of millions of tweets includes all kinds of information. Tweets are short messages of up to 140 characters. Tweets can provide a set of unique perspectives reflecting the points of



view of users who are interested in or participating in an event. In unplanned events, Twitter users sometimes spread news prior to the traditional news media (Kwak, Lee, Park, & Moon, 2010; Sakaki, Okazaki, & Matsuo, 2010). For planned events, Twitter users often post messages in anticipation of the event. In this manner, Twitter users take on the role of social sensors (Sakaki, Okazaki, & Matsuo, 2010) with the network serving as a channel for breaking news alerts and subsequently streaming real-time data as events unfold. For example during the Arab Spring, Twitter connected Western and Arab individuals to protest participants bringing unique and unfiltered content. There was also evidence foreign governments directly monitored social media sites to supplement their limited knowledge of what was actually occurring during the Arab Spring protests (Becker, Naaman, & Gravano, 2011).

From an analysis perspective, identifying low-grade intelligence on Twitter is a challenging problem due to the diversity of message content and writing style, as well as, the immense scale of the data (i.e., half a billion tweets a day; Costolo, 2012). Twitter users post messages with a variety of content types including personal updates, calls for participation, warnings about violence and threatening situations, and various other bits of information (Naaman, Boase, & Lai, 2010). In addition, Twitter messages by design, contain little textual information and are often written informally with poor spelling, grammar and sentence structure (Becker, Naaman, & Gravano, 2011). To address these challenges, new text analytic approaches such as Contour's frame-based semantic search capability are needed for extracting low-grade intelligence from social media message streams.

For demonstration purposes, a large set of tweets collected during the Arab Spring protest was imported into Contour's semantic search engine. The tweet set consists of approximately 7.3 million tweets collected between February 01 to February 19, 2011 by the Blender Cross-source Information Extraction Laboratory at the City University of New York (CUNY) Graduate Center under the direction of Dr. Heng Ji (Zubiaga, Ji, and Knight, 2013). This date range covers the heart of the rebellion from a few days after the first coordinated mass protests against Mubarak on January 25, including Mubarak's resignation on February 11, and the rejection of a power transfer to civilian administration by the Egyptian military on February 13.

We have developed a simplistic use case simulating a low-grade intelligence collection effort involving Contour's semantic search capabilities applied to the Arab Spring tweet dataset. The simulated user is an analyst requiring information about events that may be brewing nearby or within an area of operations. This use case will employ several different types of semantic searches performed on the tweets: a Concept search which is the broadest type of search returning instances of the selected Concept (FrameNet frame); a combined Concept and Role search returning instances of the Concept and Role (FrameNet frame and frame element); and the most targeted type of search, a Concept, Role, and Keyword search identifying the specified Keyword within the semantic role. For these searches either the filtering widgets or the Semantic Search box can be used to formulate queries.

The use case begins as an analyst receives an external report about a protest occurring in Egypt possibly near a Forward Operating Base (FOB). A captured stream of tweets is imported into Contour's semantic search engine using the Import Source and Content Type widgets. After import, a broad semantic search is performed using Contour's Concept filtering widget by selecting the *Protest* frame (Figure 4). The Protest Concept filters the 7,397,115 total tweets down to 73,328 tweets relevant to the Protest frame. Contour's results window displays 25 tweets per page. Figure 5 shows three example tweets from the first page of results using the Protest filter.



Figure 4. The Concept filtering widget with the Protest frame selected (top). Once the Concept is selected the Role filtering widget displays that frame's elements and the number of tweets containing the specific Concept and Role in parentheses (bottom).



Figure 5. Example tweets resulting from a search using the Concept filtering widget with the Protest Concept selected.

A quick glance at the first page of tweets informs the analyst that an Egyptian protest was triggered by the fall of leader Zine El-Abidine Ben Ali in Tunisia. The Egyptians appear to be protesting against rising prices, poverty, unemployment, and Mubarak's authoritarian regime. The protests are occurring somewhere in Cairo.

There are four frame elements associated with the Protest frame (Place, Degree, Action, Descriptor). Selecting the Place Role will give more information about the location of the protests. The analyst uses the Place Role filter reducing the number of relevant tweets to 7730. Figure 6 shows three example tweets from the first page of search results using the Protest Concept and Place Role filtering.

**Egypt: Political Energy Powers Exhausted Protesters** (Protest) Place  
 CNT-APP10-45175 - Friday, 2/8/2013 4:04pm - Unknown  
 Yet despite these fears of economic uncertainty, **protesters at Tahrir** remain steadfast, calling for million-strong protests every Sunday, Tuesday and Friday until their demands for the president's ouster are met.

**Cairo struggles for normalcy in face of protests Leaders look to wait out opposition** (Protest) DegreePlace  
 CNT-APP10-45176 - Friday, 2/8/2013 4:04pm - Unknown  
 Cairo's chronic traffic jams returned yesterday as the city began to adapt to the **sprawling protests in Tahrir Square** as well as the tanks, armored personnel carriers, and soldiers who continued to block some streets.

**The Battle for Tahrir Square Tuesday, Feb. 1 Hunder** (Protest) ActionPlace  
 CNT-APP10-45180 - Friday, 2/8/2013 4:04pm - Unknown  
 6 Defying government calls for a return to normalcy, **more than 100,000** turned out to **protest in the square**.

Figure 6. Example tweets resulting from a search using the Concept filtering widget with the Protest Concept and the Role filtering widget with the Place Role selected.

From these tweets the analyst identifies the primary location of the protests—Tahrir Square. By descriptions such as “sprawling” and “more than 100,000” contained in the tweets, the protests appear large in size involving many citizens. One of the tweets refers to “tanks, armored personnel carriers, and soldiers” indicating Egyptian government forces may be taking action against the protestors. Cairo’s local roadways are being affected by the protest activities. The analyst believes this may have an impact on planned patrol routes within the city and especially near the Tahrir Square area.

To verify the extensiveness of the protests the analyst selects the Degree Role filter reducing the number of relevant tweets to 5860 and confirming the protests are widespread. Figure 7 shows two example tweets resulting from the Protest and Degree filtering.

**World The Egyptian protests against lack of work,** (Protest) Degree  
 CNT-APP10-45182 - Friday, 2/8/2013 4:04pm - Unknown  
 Police and demonstrators fought running battles on the streets of Cairo on Friday in a fourth day of **unprecedented protests** by tens of thousands of Egyptians demanding an end to President Hosni Mubarak's three-decade rule.

**Souderton travel agent keeping eye on weather, Egypt** (Protest) Degree  
 CNT-APP2-35364 - Friday, 2/8/2013 3:54pm - Unknown  
 At the same time, though, she's keeping an eye on what's happening in Egypt, where **widespread protests** have led to curfews, violence, a shutdown of communications and the announcement by President Hosni Mubarak that after almost 30 years in office, he will

Figure 7. Example tweets resulting from a search using the Protest Concept filter and the Degree Role filter.

The resulting tweets also reference violence and possible communication disruptions. To further investigate the level of violent activities associated with the protests, the analyst selects a Keyword filter of “violence” combined with the Protest Concept filtering (Figure 8). The Role filters associated with the Protest Concept do not include “violence” or anything similar; therefore, a Keyword filter must be used. The word “violence” is typed into the Keyword text box. Figure 9 shows several tweets resulting from the Protest Concept and violence Keyword filters.

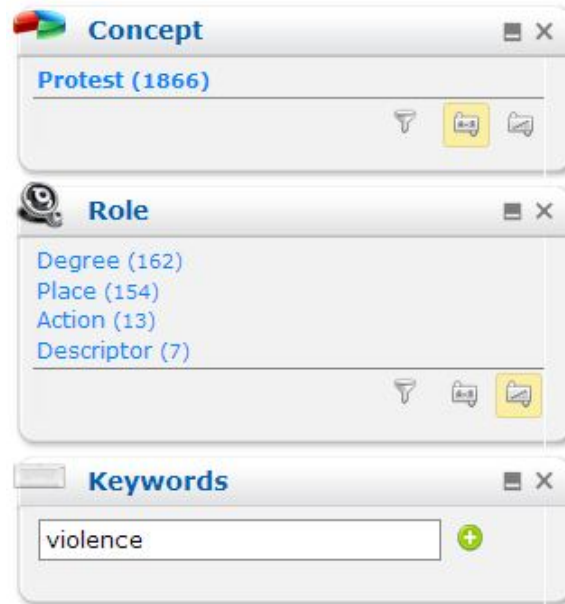


Figure 8. The Concept filtering widget with the Protest frame selected (top) and the Keyword filter “violence” typed into the Keywords widget (bottom).

**Friday, February 4, 2011 Egyptian Protests--Departing from Cairo, Accepting Naivety**
(Protest)

CNT-APP5-29488 - Friday, 2/8/2013 2:38pm - Unknown  
 The pro-government supporters, who have been in the street over the past few days, bringing with them the horrific violence that has consumed Tahrir Square are largely not real protesters, but hired vigilantes and members of the ruling party and the polic

**Published on Wednesday, February 9, 2011 by CommonDreams.org**
(Protest)

CNT-APP6-39920 - Friday, 2/8/2013 2:38pm - Unknown  
 Palestinian Authority/Hamas The Palestinian Authority's police used violence against peaceful demonstrators during a rally in Ramallah on February 2, 2011, to support the protesters in Egypt.

**Obituaries This site uses cookies. By continuing t**
(Protest)

CNT-APP9-40731 - Friday, 2/8/2013 2:38pm - Unknown  
 The protests that started with a few thousand people on January 25th escalated to a thrilling climax on February 1st, when hundreds of thousands assembled in Cairo's Tahrir Square to demand the removal of Hosni Mubarak, and then deteriorated into violence

**Middle East and Iran 0522 GMT: We can now confirm**
(Protest)

CNT-APP3-44343 - Friday, 2/8/2013 2:38pm - Unknown  
 Hosni Mubarak begins speaking, talking about how peaceful protest regretfully turned to violence, looting, pillaging, and assault on private and public property including storm of some diplomatic missions.

Figure 9. Example tweets resulting from a search using the Protest Concept filter combined with a “violence” Keyword filter.

The tweets returned by the Protest Concept and violence Keyword search indicate the protest may have started peacefully but now has turned violent with looting and damage to private and public property. The Egyptian government is utilizing equipment and tactics to control this violence and disperse the protesters. This will further disrupt traffic and impact any ground patrol units planning on entering the area.



In addition to restricted ground travel, the analyst is concerned about communications limitations in the Cairo area because the tweet content from a previous search mentioned “a shutdown of communications” (Figure 7). Again, the analyst uses a Keyword filter in conjunction with the Protest Concept filter. Figure 10 shows some example tweets returned from a search using the Keyword “communication”. These tweets inform the analyst the government has shut down a major internet provider and cell phone networks in the hopes of dispersing the protestors. Also, Egyptian government soldiers appear to be communicating by CB radio. The analyst believes the information about the communications blackout should be sent up the chain of command.

The tweet mentioning the use of CB radios by “Thugs” implies government soldiers are “attacking protesters.” Additional information is needed to determine the level of violence against civilians. In this case, the analyst uses the Semantic Search box for a more targeted search approach. The Concept *Killing*, and the Role *Victim* is selected from dropdown lists populated by the FrameNet ontology of frames and associated frame elements. When using the Semantic Search box, a Keyword is required with optional wildcard characters. The Keyword of *\*Cairo\** is entered because the analyst is interested in killings linked to the protest activities occurring in Cairo. Figure 11 shows the search criteria (Concept, Role, and Keyword) entered in the Semantic Search box. Figure 12 displays some of the tweets returned by this search where the Killing Concept identified sentences containing words such as shooting, killed, and assassination (blue and yellow highlighted text). The green highlighted text corresponds to the use of the *\*Cairo\** keyword.

**Egyptian government tightens grip ahead of protests**
(Protest)

CNT-APP10-4202 - Wednesday, 2/6/2013 11:48am - Unknown

Egyptian government tightens grip ahead of **protests**. By the CNN Wire Staff Gunfire amid protests in Cairo STORY HIGHLIGHTS Major Internet provider shut down Mobile phone networks to be shut down Trains ordered stopped indefinitely Internet (CNN) -- The emb

**Egypt News.Net Tuesday 15th February, 2011 Iranian**
(Protest)

CNT-APP10-4182 - Wednesday, 2/6/2013 11:47am - Unknown

Authorities have also blocked websites and satellite news channels in an effort to disrupt communications that could help organise **protests**.

**12,000 notes}} The Egyptian Revolution of 2011 ( t**
(Protest)

CNT-APP3-4009 - Wednesday, 2/6/2013 11:47am - Unknown

The Egyptian government shut down the Internet to most of Egypt during the recent **protests** in order to limit communication between protest groups.

**\u267a @helkomy: Thu**
(Protest)

CNT-APP8-74419 - Friday, 2/15/2013 9:19am - English

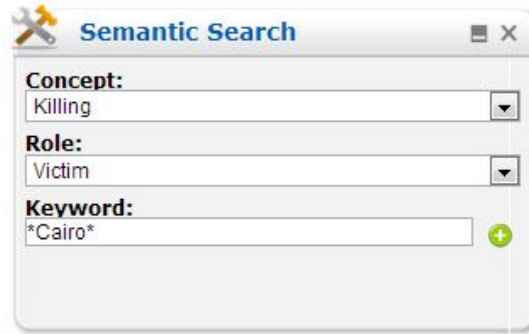
Thugs are using CB radio to communicate to keep **attacking protesters** #tahrir cc ioerror - can we build 27/40Mhz jammers?

**Liberation Square Pr**
(Protest)

CNT-APP10-39515 - Friday, 2/8/2013 9:37am - English

Liberation **Square** **Protests** are Still on for Tomorrow despite Government Communications BlackOut| #MarchOfAMillion #Egypt #Cairo

Figure 10. Example tweets resulting from a search using the Protest Concept filter combined with a “communication” Keyword filter.



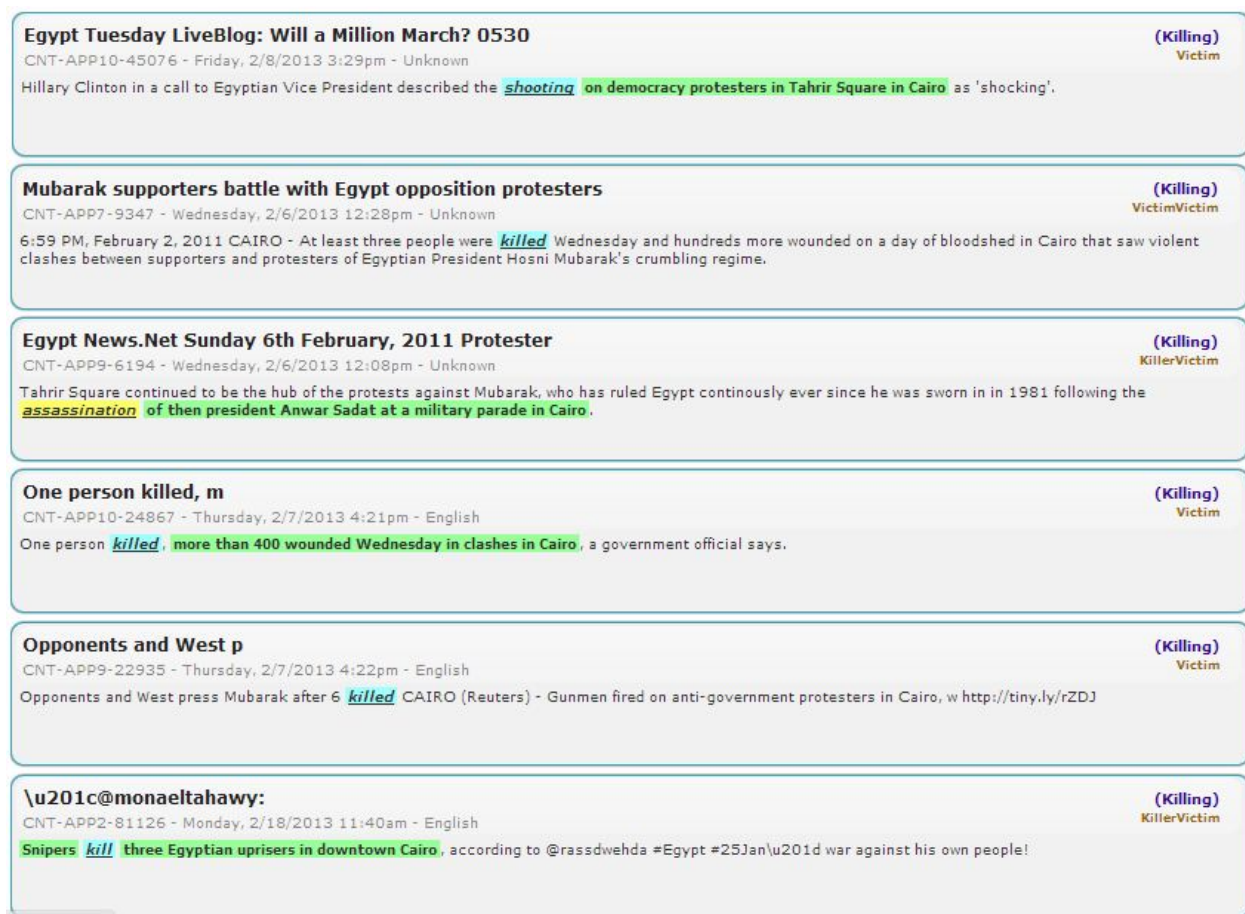
**Semantic Search**

**Concept:**  
Killing

**Role:**  
Victim

**Keyword:**  
\*Cairo\*

Figure 11. The Semantic Search box with the Killing Concept, Victim Role, and the Keyword \*Cairo\* entered as search criteria.



**Egypt Tuesday LiveBlog: Will a Million March? 0530** (Killing) Victim  
 CNT-APP10-45076 - Friday, 2/8/2013 3:29pm - Unknown  
 Hillary Clinton in a call to Egyptian Vice President described the shooting on democracy protesters in Tahrir Square in Cairo as 'shocking'.

**Mubarak supporters battle with Egypt opposition protesters** (Killing) VictimVictim  
 CNT-APP7-9347 - Wednesday, 2/6/2013 12:28pm - Unknown  
 6:59 PM, February 2, 2011 CAIRO - At least three people were killed Wednesday and hundreds more wounded on a day of bloodshed in Cairo that saw violent clashes between supporters and protesters of Egyptian President Hosni Mubarak's crumbling regime.

**Egypt News.Net Sunday 6th February, 2011 Protester** (Killing) KillerVictim  
 CNT-APP9-6194 - Wednesday, 2/6/2013 12:08pm - Unknown  
 Tahrir Square continued to be the hub of the protests against Mubarak, who has ruled Egypt continuously ever since he was sworn in in 1981 following the assassination of then president Anwar Sadat at a military parade in Cairo.

**One person killed, m** (Killing) Victim  
 CNT-APP10-24867 - Thursday, 2/7/2013 4:21pm - English  
 One person killed, more than 400 wounded Wednesday in clashes in Cairo, a government official says.

**Opponents and West p** (Killing) Victim  
 CNT-APP9-22935 - Thursday, 2/7/2013 4:22pm - English  
 Opponents and West press Mubarak after 6 killed CAIRO (Reuters) - Gunmen fired on anti-government protesters in Cairo, w http://tiny.ly/rZDJ

**\u201c@monaeltahawy:** (Killing) KillerVictim  
 CNT-APP2-81126 - Monday, 2/18/2013 11:40am - English  
 Snipers kill three Egyptian uprisers in downtown Cairo, according to @rassdwehda #Egypt #25Jan\u201d war against his own people!

Figure 12. Example tweets resulting from using the Semantic Search box with search criteria Killing Concept, Victim Role and the Keyword \*Cairo\*.

This use case demonstrated how an analyst could collect low-grade intelligence information from a large stream of social media data using Contour's semantic search interface. With the frame-based filtering widgets (Concept and Role) and an optional keyword to further narrow the search, the analyst was able to quickly and concisely construct a series of queries using a mouse and dropdown menus. These queries



were used to iteratively search through the content of over seven million tweets for relevant bits of information pertaining to events occurring within an area of operations.

The low-grade information extracted from the tweets, when added together, forms an operational picture of widespread protests in the area of Tahrir Square, Cairo. The protests which began as peaceful demonstrations against Mubarak and his government rapidly escalated to violent clashes with the Egyptian military. The analyst learned transportation routes were blocked and communication services shut down. The Egyptian military communicating via CB radios and using tanks were firing on protestors resulting in many civilian casualties. This information can complement high-grade intelligence supporting command level decision makers in carrying out ongoing missions and mitigating emergent threats.

## Discussion

The use case in the previous section, although simplistic in nature, portrays how an analyst could search a large stream of tweets for low-grade information that when added together produces actionable intelligence within a collection resource deprived operational environment. Two of the most important concerns of the analyst would be to gather information about what events are occurring or may occur within a specific geographic region such as an area of operations or forward operating base, and how these events might affect missions such as ground patrol units or forward communications. The use case simulated the construction of several search queries at different levels of complexity to extract useful information from seven million tweets reporting the rapidly deteriorating situation in Tahrir Square.

Contour's widget-based search filtering system offers many advantages. Queries at three different levels of complexity can be constructed using a mouse with minimal text input. Query construction is intuitive with dropdown lists displaying the available FrameNet Concepts (frames) and associated Roles (frame elements) for mouse-click selection, and an optional Keyword text entry widget to narrow the query focus further. In addition, the designed simplicity of the main window of the user interface lends itself to mobile computing devices where display screen size is restricted. Other capabilities of the semantic search interface not covered in the use case include searching by pre-defined date range filters, and a text box for more traditional manual query entry using Concepts, Roles, Keywords including wildcard characters and logical operators.

There are many challenges and stresses associated with the collection and analysis of high-grade intelligence, for example, the lack of network security to access classified reports or unsafe conditions to observe and document potentially threatening activities. In these types of data processing environments, Contour's point and click query construction enables the collection of low-grade intelligence information from easily accessible social media sources by inexperienced analysts in support of a common operating picture.

Contour's semantic search technology is in the prototype stage of development. There are two areas for improvement—extending the FrameNet ontology, and integrating real-time processing of social media content. Contour's Concepts and Roles widgets are populated by the FrameNet ontology. The FrameNet project represents the culmination of decades' worth of research on how to depict knowledge for language processing systems. FrameNet, primarily funded by the National Science Foundation and housed at the International Computer Science Institute in Berkeley, California, was designed as a general ontology to model the meaning of text. Intelligence analysts draw upon data from many different types of sources including command and control and mission support; therefore, ontologies for semantic search systems must be able to accommodate a wide variety of Concepts and Roles. Ontology gaps can become problematic when attempting to analyze text within dynamic military environments. Because intelligence analysts frequently need to understand and adapt to new situations, methods of automatically expanding FrameNet's ontology would improve the future capabilities of Contour's semantic search engine.

Currently, the social media collection and pre-processing, and the semantic search analysis are a two-step process utilizing different technology platforms. An external collection mechanism is used to capture streams of social media content such as tweets. Then the captured text messages are pre-processed (i.e., SLR), and imported into Contour's semantic search engine. One improvement to this process to better utilize the dynamic nature of social media is to seamlessly ingest, pre-process, and analyze the message streams for semantic search. Identifying events as they occur or even anticipating what kind of event will happen, and where, based on social media user-contributed messages, would be a valuable intelligence collection asset and a future area of development for Contour.

## Conclusion

Social media users produce millions of messages a day sometimes live-tweeting about critical events such as natural disasters or political transitions. While much of the content of social media is unstructured and non-informative, advanced analytic technologies such as Contour's FrameNet-based semantic search engine can identify small bits of low-grade intelligence information present within social media streams. This low-grade information when added together can produce actionable intelligence within a collection resource deprived operational environment. Contour's semantic search process maps words and phrases in social media content to FrameNet frames and frame elements using Concept and Role filtering widgets. This widget interface allows three levels of complexity in query formulation using intuitive dropdown menus and optional keyword text entry.

A use case demonstrated Contour's primary semantic search capabilities in the context of an analyst tasked with gathering information about events occurring within a specific geographic region such as a forward operating base. The use case simulated the construction of several search queries at different levels of complexity to extract useful information from over seven million tweets reporting the rapidly deteriorating security situation in Tahrir Square during the Egypt uprisings.

## Acknowledgments

The authors acknowledge the developers of the Contour semantic search technology, Decisive Analytics Corporation (DAC), and their supporting scientists: Jack Davenport, Peter David, and Scott Martin. This technology is currently under development with funding from SBIR topics on text analytics. The authors also acknowledge the Blender Cross-source Information Extraction Laboratory at the City University of New York (CUNY) Graduate Center directed by Dr. Heng Ji and graduate student Hongzhao Huang for assisting in acquiring the Arab Spring Twitter dataset as part of research performed under the U.S. Army Research Laboratory Network Science Collaborative Technology Alliance (NS CTA).

## References

Becker, H., Naaman, M., & Gravano, L. (2011). Beyond trending topics: Real-world event identification on twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain, July 17-21.

Costolo, D. (2012). Retrieved April 22, 2013, from <http://www.complex.com/tech/2012/10/twitter-ceo-dick-costolo-reveals-staggering-number-of-tweets-per-day>

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280, 20-32.

- Gildea, D., & Jurafsky, D. (2002). Automatic Labeling of Semantic Roles. *Computational Linguistics* 28(3), 245-288.
- Kitson, F. (1971). *Low intensity operations: Subversion, insurgency, peacekeeping*. London: Faber and Faber.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19<sup>th</sup> International World Wide Web Conference*. Raleigh, North Carolina, April 26-30.
- Naaman, M., Boase, J., & Lai, C. H. (2010). Is it really about me?: Message content in social awareness streams. In *Proceeding of the 14<sup>th</sup> International Conference on Computer Support Cooperative Work*, Savannah, Georgia, February 6-10.
- Nextgov. (2012). Retrieved March 3, 2013, from <http://www.nextgov.com/mobile/2012/07/social-media-plays-growing-role-open-source-intelligence-experts-say/57143/>
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. J., & Scheffczyk, J. (2006). FrameNet II: Extended Theory and Practice. Retrieved from the FrameNet website: <http://framenet.icsi.berkeley.edu>
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19<sup>th</sup> International World Wide Web Conference*. Raleigh, North Carolina, April 26-30.
- Zubiaga, A., Ji, H., & Knight, K. (2013). Curating and contextualizing Twitter stories to assist with social newsgathering. In *Proceeding of the International Conference on Intelligent User Interfaces*. Madeira, Portugal, January 13-16.



**US ARMY  
RDECOM**

U.S. Army Research, Development and Engineering Command

# Accelerating Exploitation of Low-grade Intelligence Through Semantic Text Processing of Social Media

# ARL

**TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.**

Dr. Sue E. Kase & John Dumer

U.S. Army Research Laboratory

Computational Information Sciences Directorate

Tactical Information Fusion Branch

18<sup>th</sup> ICCRTS, June 2013

# SOCIAL MEDIA CONTENT

Increasing usage of  
mobile devices



Facilitates information  
sharing from anywhere  
at any time



## Social Media

Unprecedented volumes of real-time  
information contributed by large  
communities of social media users

Is social media potentially valuable to the  
intelligence community?

How can social media be searched for information  
relevant to national security?



# WHAT IF NO TRADITIONAL SOURCES OF INTELLIGENCE AVAILABLE



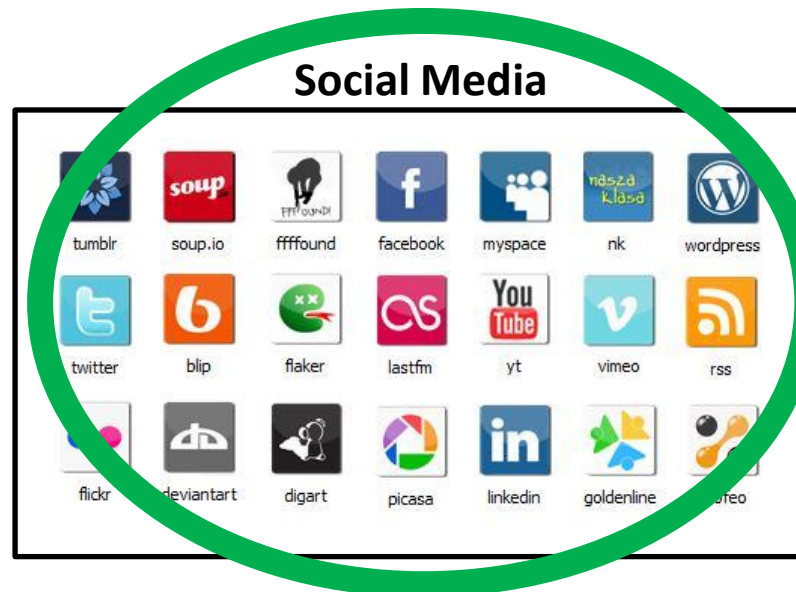
*Constrained resource  
allocation environment*



*Challenges and risks  
with direct intelligence  
data collection*

Intelligence collection  
relying on the eyes and  
ears of the general  
population

Citizen-authored  
information



A source of 'low-grade'  
intelligence information if  
search and extraction  
technologies are  
developed



# FRAME-BASED SEMANTIC SEARCH OF SOCIAL MEDIA

## CHALLENGE

Social media content is dynamic, 'unstructured' containing hashtags, urls, emoticons, replies and mentions, and little useful information



atnussan @atnussan

12h

#Afghan #interpreter's family killed by Taliban near Kandahar | Toronto Star [bit.ly/187mx3y](http://bit.ly/187mx3y) cc @David\_Cameron  
Expand



CrossFit @CrossFit

21 May

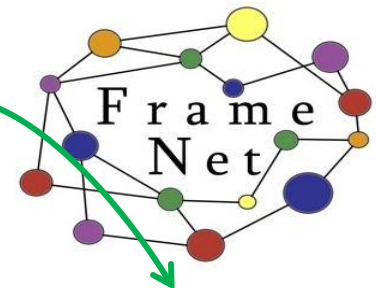
RT @USAFCENT: Members of Kandahar @CrossFit honored four fallen Airmen with a workout to remember. [on.fb.me/118LT8j](https://on.fb.me/118LT8j) @usairforce

## SOLUTION

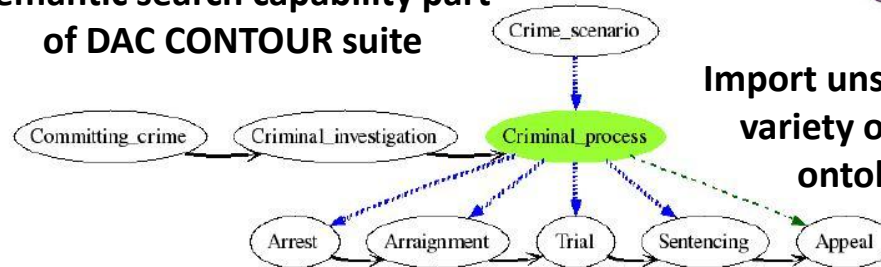


Distill social media for relevant information using frame-based semantic modeling

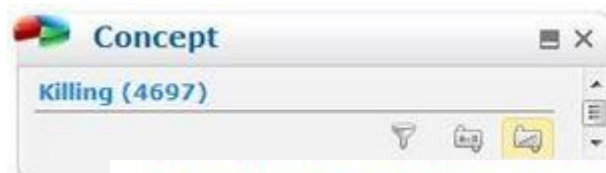
Semantic search capability part of DAC CONTOUR suite



Import unstructured text from variety of sources, map to ontology of frames



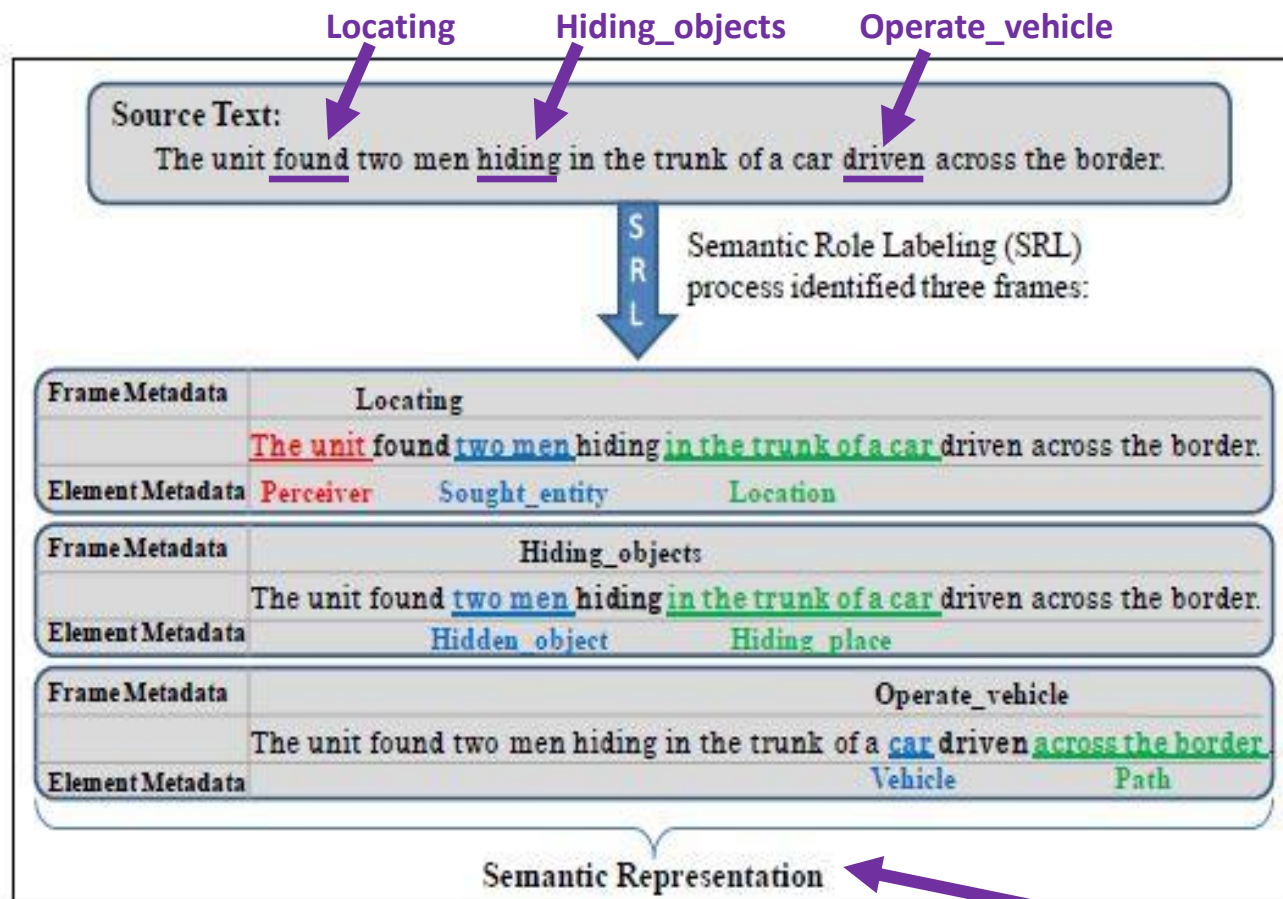
Construct queries using frame (Concept) and frame element (Role) filtering widgets



# FrameNet-based SRL

Frame Semantics theory (Fillmore & Colleagues, 1976)  
Semantic Role Labeling (SRL) (Gildea & Jurafsky, 2002)

Berkeley FrameNet lexical database  
<https://framenet.icsi.berkeley.edu/fndrupal/>



**Frame:** a schematic representation of a situation involving various participants, props, and other conceptual roles

3 frames in target sentence  
Locating ('found')  
Hiding\_objects ('hiding')  
Operate\_vehicle ('driven')

Locating Frame has  
Frame Elements  
Perceiver  
Sought\_entity  
Location

**Frame Element :** frame-specific defined semantic role that is the basic unit of a frame

## Comparison with WordNet and ontologies

Multiple annotated examples (~20) of each LU.  
Examples taken from naturalistic corpora.  
LU linked to semantic frame similar to thesaurus.

All combinatorial possibilities of the LU  
Analysis proceeds frame by frame  
Network of relations between frames

**Builds a semantic model of the text**

# FrameNet-based QUERY BUILDING INTERFACE

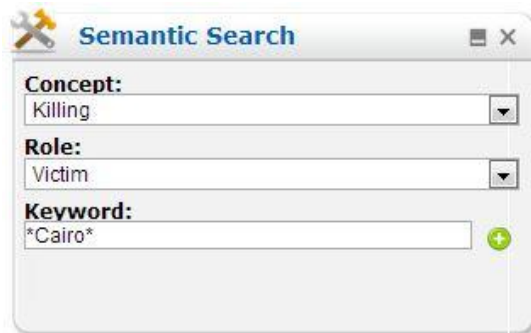
FrameNet-based SRL → Semantic model of the text → Concept + Role filtering widgets

*Mapping between  
Frame Concepts + Frame Element Roles  
and the target text*

*Select Frame from  
dropdown menu in  
Concept widget*

*Then Select Frame  
Element from  
dropdown menu in  
Role widget*

*Alternative Example: Using the  
Semantic Search widget*



**Semantic Search**

**Concept:**  
Killing

**Role:**  
Victim

**Keyword:**  
\*Cairo\*

*Killing Frame with  
Victim Frame Element  
selected  
Keyword \*Cairo\* entered  
for targeted search*

*Example: Killing Frame selected  
Frame Elements of Killing Frame  
displayed in Role widget*



**Concept**

Killing (4697)

**Role**

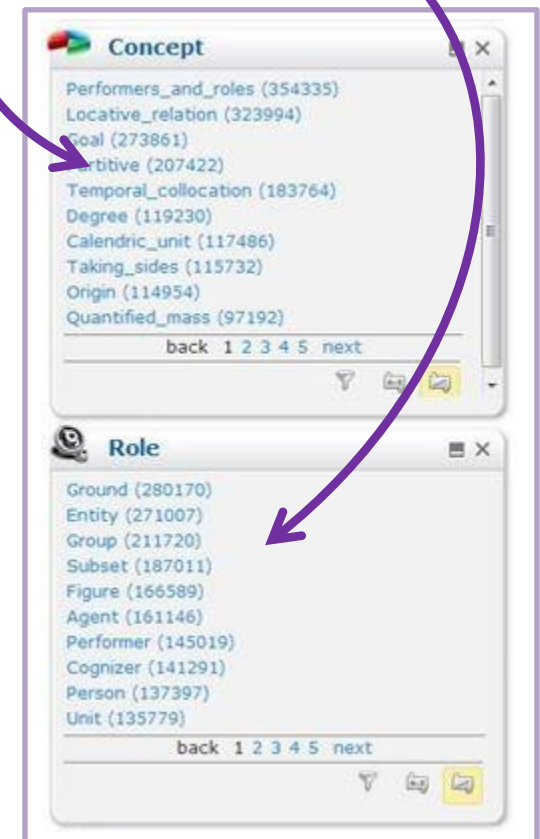
Victim (4697)

Killer (1575)

Cause (418)

Time (13)

*Then Select Frame Element*



**Concept**

Performers\_and\_roles (354335)

Locative\_relation (323994)

Goal (273861)

Partitive (207422)

Temporal\_collocation (183764)

Degree (119230)

Calendric\_unit (117486)

Taking\_sides (115732)

Origin (114954)

Quantified\_mass (97192)

back 1 2 3 4 5 next

**Role**

Ground (280170)

Entity (271007)

Group (211720)

Subset (187011)

Figure (166589)

Agent (161146)

Performer (145019)

Cognizer (141291)

Person (137397)

Unit (135779)

back 1 2 3 4 5 next



# USE CASE: SEMANTIC SEARCH OF TWEETS



Twitter: streaming real-time data, immense scale

Twitter users: act as social sensors, post in anticipation of events, report breaking news

Twitter messages: little textual information, informally written, embedded non-text characters

Tweet content types: personal updates, calls for participation, warnings about threatening situations

Traditional sources of intelligence information are not currently available

Use CONTOUR's FrameNet-based semantic search on social media

Demonstrate how CONTOUR's query interface can be used to identify low-grade intelligence information



Tweets from Arab Spring protest

Data set: 7.3 million tweets

Collected between Feb 01 – 19, 2011

NS CTA INARC collaboration:

Blender Cross-source Information

Extraction Laboratory

City University of New York (CUNY)

Dr. Heng Ji

## SEMANTIC SEARCH OF TWEETS

[ DEMO or MOVIE ]

An analyst receives an external report about a protest occurring in Egypt possibly near a Forward Operating Base (FOB)

The analyst's chain of command needs to know what is going on in that area; there is no time to collect intel

A stream of tweets is captured and imported into CONTOUR's semantic search system

## ACKNOWLEDGMENTS



SBIR Phase II  
Decisive Analytics  
Corporation

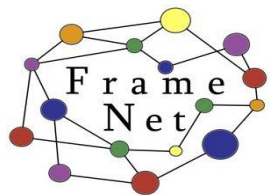


[http://www.dac.us/Capabilities/Intelligent\\_Products](http://www.dac.us/Capabilities/Intelligent_Products)



ARL NS CTA  
BLENDER Lab, CUNY  
Dr. Heng Ji

<http://nlp.cs.qc.cuny.edu/>



FrameNet Project

<https://framenet.icsi.berkeley.edu/fndrupal/>

Utilize social media to gather information about events occurring or events that may occur within a geographic region such as an area of operations

Develop new text analytics technologies such as CONTOUR's Frame-based semantic search to process social media content that is unstructured and primarily non-informative

Use Case demonstrates how an analyst can search a large stream of tweets with frame-based queries constructed from Concepts, Roles, and Keywords

Collecting low-grade information from social media - when added together - produces actionable intelligence such as determining how an event may affect missions